Candidata: Gleiciane Galdino

Vaga: Estágio em IA

Relatório Técnico: Orquestração de IA com n8n para Análise de Dados

O objetivo desse projeto é desenvolver uma Prova de Conceito (PoC) de um sistema de Business Intelligence, no qual o n8n centraliza a orquestração da migração de dados de planilhas para um banco de dados e atua como motor de uma aplicação de IA conversacional, permitindo que o usuário faça perguntas em linguagem natural e receba respostas precisas baseadas nos dados.

Obtendo a estrutura do projeto:

Ao clonar ou fazer o fork do repositório, você encontrará a seguinte estrutura de arquivos:

/Infraestrutura-Base

— Dockerfile

(Instruções para construir a imagem da API)

— docker-compose.yml (Inicia a API, o banco de dados e o n8n)

├── main.py

(Código da API em Python)

requirements.txt (Dependências do Python)

Acessando a pasta do projeto pelo terminal:

Para rodar o projeto, você precisa primeiro abrir um terminal ou PowerShell dentro da pasta que contém o arquivo docker-compose.yml.

Existem duas formas fáceis de fazer isso:

• Usando o comando cd: No terminal, digite cd (que significa "mudar diretório") seguido do caminho da sua pasta e pressione Enter. Por exemplo: cd C:\Users\SeuUsuario\Desktop\Projetos\Infraestrutura-Base

 Direto do Visual Studio Code: Se você já usa o VS Code, pode abrir o terminal diretamente na pasta do projeto. No menu superior, vá em Terminal e selecione New Terminal. O terminal já abrirá no local certo, pronto para você usar.

Como rodar o projeto:

Primeiro deve abrir o docker desktop e acessar a conta cadastrada; o próximo passo é acessar a pasta do projeto (explicado anteriormente), chamada Infraestrutura-Base, que contém o arquivo docker-compose.yml, abrir um terminal e executar o comando docker-compose up -d. O docker-compose.yml está configurado para iniciar a API, o banco de dados e o n8n de uma vez, e eles continuarão rodando em segundo plano.

O comando **docker-compose up**: vai ler o arquivo **docker-compose.yml** e iniciar todos os serviços definidos nele:

- Construindo a imagem da API;
- Baixar a imagem do banco de dados e do n8n;
- Iniciar todos os containers.

A flag **-d** significa "detach", o que faz com que os containers rodem em segundo plano, permitindo o uso contínuo do terminal.

### Ou seja:

- 1. Abrir o Docker Desktop e acessar a conta cadastrada;
- 2. Acessar a pasta do projeto Infraestrutura-Base;
- 3. Executar o comando: docker-compose up -d.

#### Como saber se deu certo:

Para verificar o status dos containers execute o comando: **docker ps**, a saída mostrará os containers **api\_container**, **postgres\_container** e **n8n\_container** listados e com o status **Up**.

### Acessando os Serviços:

API: Estará disponível na porta 8000.
 Pode testá-la acessando <a href="http://localhost:8000">http://localhost:8000</a> no seu navegador. A resposta deve ser {"status": "API está funcionando!"} como na imagem abaixo:



{"status":"API está funcionando!"}

• n8n: A instância do n8n estará disponível na porta 5678. Acesse http://localhost:5678 no seu navegador e faça login com as credenciais padrão: usuário admingle e senha admingle123.

 PostgreSQL: O banco de dados está na porta 5432. Ele só é acessível pelos outros containers. Para interagir com ele, você pode usar uma ferramenta de banco de dados ou entrar no container com docker exec -it postgres\_container psql -U admingle -d meubanco.

## Etapa 1: Estrutura do Projeto - Infraestrutura Base (Docker, Python, Postgres)

**1º Passo:** Criei o arquivo docker-compose.yml, responsável para iniciar a API, banco de dados e o n8n. A primeira instrução foi o banco de dados, contendo uma arquitetura simples apenas para suportar os dados disponibilizados.

#### Estrutura do banco e tabela:

Criei um banco de dados relacional no PostgreSQL para centralizar os dados de eventos que antes estavam dispersos em planilhas, para isso, utilizei o comando: **docker-compose exec postgres\_container psql -U admingle -d meubanco.** 

Dentro dele, criei a tabela **eventos**, que armazena informações de cada evento de forma organizada e padronizada. Usei uma tabela única, porque todos os dados possuem o mesmo contexto (eventos).

### Estrutura da tabela:



A tabela foi criada de forma flexível para garantir que todos os dados vindos das planilhas integradas pelo n8n sejam inseridos no banco sem falhas, priorizando a demonstração do fluxo automatizado em vez da validação rígida dos dados.

2° Passo: Criei o arquivo main.py que contém o código principal da API, construída com o framework FastAPI.

Nele, foram definidos os seguintes pontos:

Conexão com o banco de dados PostgreSQL, utilizando a biblioteca psycopg2;

- Definição do modelo de dados Evento com Pydantic, que garante validação dos campos recebidos;
- Implementação dos endpoints principais:
  - GET / rota de teste da API;
  - POST /eventos/ cria um novo evento;
  - GET /eventos/ lista todos os eventos cadastrados;
  - GET /eventos/{nome\_evento} consulta um evento específico;
  - PUT /eventos/{nome\_evento} atualiza um evento;
  - o DELETE /eventos/{nome evento} apaga um evento.

Esse arquivo representa a camada de acesso aos dados, permitindo que a aplicação insira, consulte, atualize e remova registros no banco.

**3° Passo:** Criei o arquivo **requirements.txt** para conter a lista de dependências necessárias, garantindo que a API funcione corretamente. O Docker utiliza esse arquivo para instalar automaticamente todas as bibliotecas no ambiente do container.

#### Inclui as bibliotecas:

- fastapi → framework web utilizado.
- uvicorn[standard] → servidor ASGI para rodar a aplicação.
- psycopg2-binary → driver para conectar ao banco PostgreSQL.
- **4° Passo:** Criei o arquivo **Dockerfile** que define como a imagem da API deve ser construída. Esse arquivo garante que a API seja reproduzida em qualquer ambiente, de forma padronizada.

# Principais etapas configuradas:

- Baseado na imagem oficial do Python 3.9 slim;
- Configuração do diretório de trabalho /app;
- Instalação das dependências a partir do requirements.txt;
- Cópia do código da API para dentro da imagem;
- Exposição da porta 8000, utilizada pelo FastAPI;
- Comando para iniciar a aplicação com Uvicorn.

**5° Passo:** Atualizei o arquivo **docker-compose.yml** para orquestrar os serviços. Nessa etapa, são iniciados dois containers, garantindo que o banco de dados e a API sejam executados de forma integrada, o que simplifica o gerenciamento:

# 1. Postgres

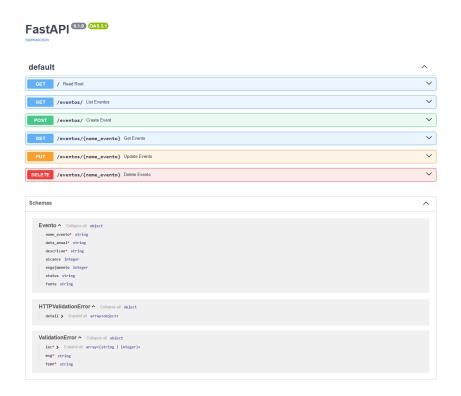
- Banco de dados relacional para armazenar os eventos;
- o Configurado com usuário, senha e nome do banco via variáveis de ambiente;
- Persistência dos dados através de volume mapeado.

## 2. API (FastAPI)

- Contêiner construído a partir do Dockerfile;
- o Conectado ao banco de dados via variáveis de ambiente;
- Porta 8000 exposta para acessar os endpoints da API;
- Dependente do serviço do banco (depends\_on).

A imagem abaixo mostra a documentação automática do FastAPI:

Acessando pelo link <a href="http://localhost:8000">http://localhost:8000</a> aparece aparecer {"status": "API está funcionando!"} e pelo <a href="http://localhost:8000/docs">http://localhost:8000/docs</a> aparece a documentação automática do FastAPI:



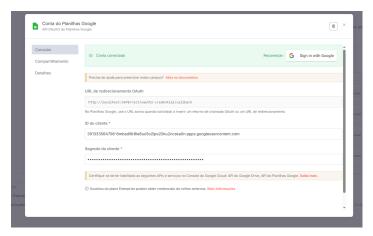
# Etapa 2: Workflow de Migração de Dados (ETL com n8n):

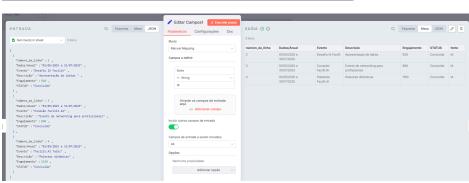
1º Passo: Atualizei o arquivo docker-compose.yml adicionando o n8n, que agora inclui os três serviços (PostgreSQL, API e n8n), permitindo que eles se comuniquem. Após a execução do arquivo docker-compose.yml com comando docker-compose up -d --build, acessei o endereço http://localhost:5678 com login e senha que criei, começando a usar o workflow para a migração de dados.

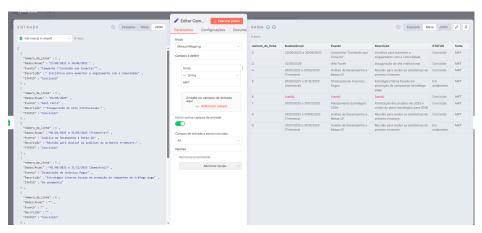
2º Passo: Estabeleci um fluxo de trabalho no n8n e o processo começou de forma manual, executando o fluxo. Criei uma conta no Google Sheets e conectei ao n8n. Autorizei o acesso com minhas credenciais seguindo o passo a passo que o próprio n8n mostrou. A próxima etapa foi configurar a leitura de três planilhas distintas usando nós Google Sheets. Cada planilha contém informações que precisei unificar em uma única estrutura. Para garantir a consistência dos dados, criei etapas intermediárias de edição de campos, onde padronizei colunas, ajustei formatos (como datas e textos) e corrigi divergências entre as diferentes fontes.

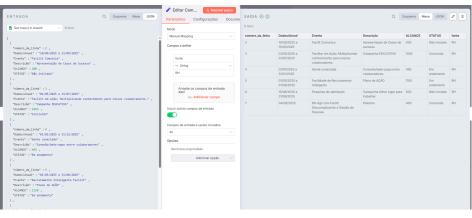
### Ou seja:

- Criei um fluxo de trabalho que começa manualmente, quando executo o fluxo;
- Configurei uma conta no Google Sheets e conectei ao n8n;
- Autorizei o acesso com minhas credenciais, seguindo o passo a passo do próprio n8n:
- Configurei a leitura de três planilhas distintas usando nós do Google Sheets;
- Cada planilha continha informações que eu precisava unificar em uma estrutura única:
- Para garantir a consistência dos dados, criei etapas intermediárias de edição de campos, onde:
  - o Padronizei colunas;
  - Ajustei formatos (como datas e textos);
  - Corrigi divergências entre as diferentes fontes.

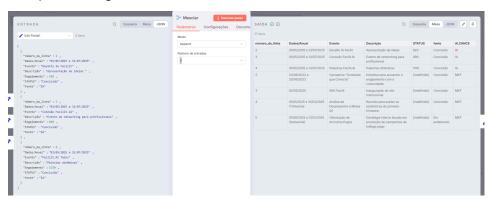




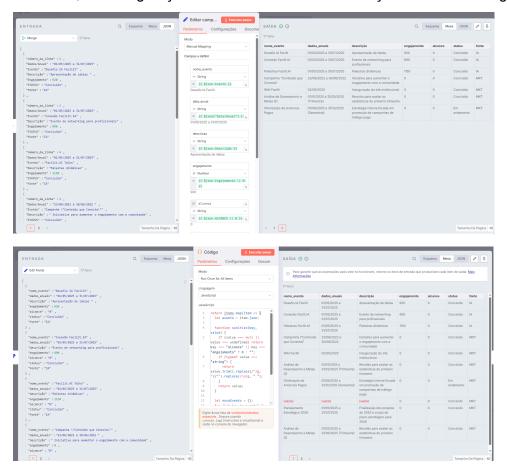




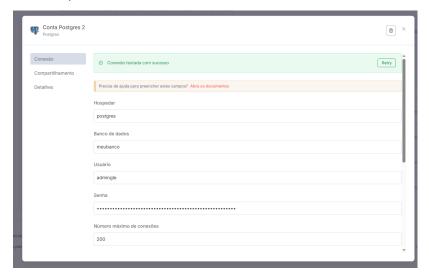
**3° Passo:** Após esse ajuste, implementei um nó de **mesclagem** para consolidar os registros em um único conjunto de dados. Dessa forma, os itens provenientes das **três planilhas** foram reunidos em uma coleção final, eliminando redundâncias e preparando tudo para a carga no banco.

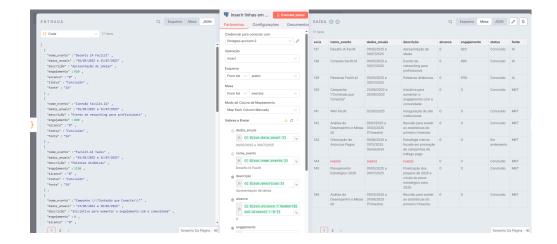


**4º Passo:** Para refinar ainda mais as informações, adicionei uma etapa de **edição de campos**, seguida por um nó **code** customizado, que foi utilizado para aplicar regras adicionais, como a geração de identificadores e a validação de valores obrigatórios.

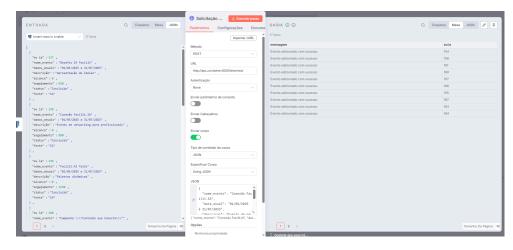


**5º Passo:** Com os dados devidamente tratados, configurei a etapa de inserção em uma tabela do **PostgreSQL**. Criei a credencial do Postgres no n8n, conectando o banco chamado meubanco. Para isso, informei o host do serviço, o **usuário** e a **senha** configurados no **docker-compose.yml**. Após isso, testei a conexão e validei que estava funcionando. Nessa fase, os registros já entram padronizados e estruturados conforme o modelo que defini anteriormente no banco.





**6° Passo:** Por fim, adicionei uma chamada **HTTP** do tipo **POST** para a **API** que construí em FastAPI. Essa etapa permite que os dados inseridos no banco também sejam transmitidos para a API, servindo tanto para validação como para a integração com outras camadas da solução.



e quando acessei endereço <a href="http://localhost:8000/eventos/">http://localhost:8000/eventos/</a> mostrou os dados salvos no banco de dados .

```
Estilos de formateaco 

"alcance": 0, "astatus": Transguração uo Site Intitucionai:, "alcance": 0, "astatus": Concluído", "fonte": "MKI"

"nome evento": "Análise de Desempenho e Metas Q1", "data anual": "01/01/2025 à 31/03/2025 (Trimestral)", "data anual": "01/01/2025 à 31/03/2025 (Trimestral)", "alcance": 0, "status": Concluído", "fonte": "MKI"

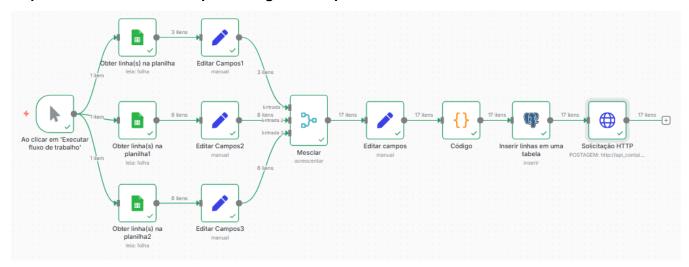
"nome evento": "Otimização de Anúncios Pagos", "data anual": "01/08/2025 à 31/12/2025 (Semestral)", "descricao": "Estratégia interna focada em otimizar campanhas de tráfego pago", "alcance": 0, "status": "Es andemento", "fonte": "MKI"

"nome evento": "NKI"

"nome evento": "", "data anual": "01/08/2025 à 31/07/2025", "data anual": "01/08/2025 à 31/07/2025", "data anual": "01/08/2025 à 31/07/2025", "descricao": "finalização dos projetos de 2025 e crição de plano estratégico para 2026", "alcance": 0, "engajamento": "Milise de Desempenho e Metas Q1", "data anual": "01/08/2025 à 31/08/2025 (Trimestral)", "descricao": "Finalização dos projetos de 2025 e crição de plano estratégico para 2026", "alcance": 0, "engajamento": 0, "engajamento": 0, "engajamento": 0, "engajamento": "MKI"

"nome, evento": "Nalise de Desempenho e Metas Q1", "data, anual": "01/08/2025 à 31/08/2025 (Trimestral)", "descricao": "Reunião para avaliar as métricas do primeiro trimestre.", "alcance": 0, "engajamento": 0,
```

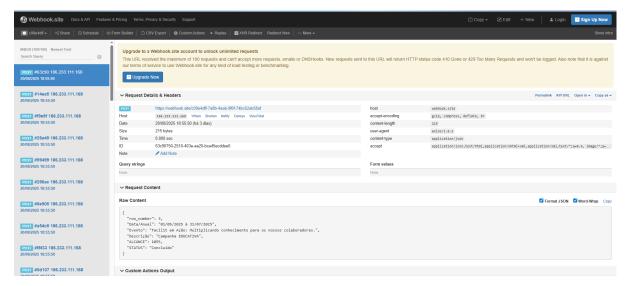
# Arquitetura do fluxo criado para a segunda etapa do desafio:



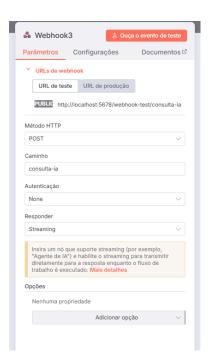
Etapa 3: Workflow de Consulta com IA (O Cérebro do Projeto):

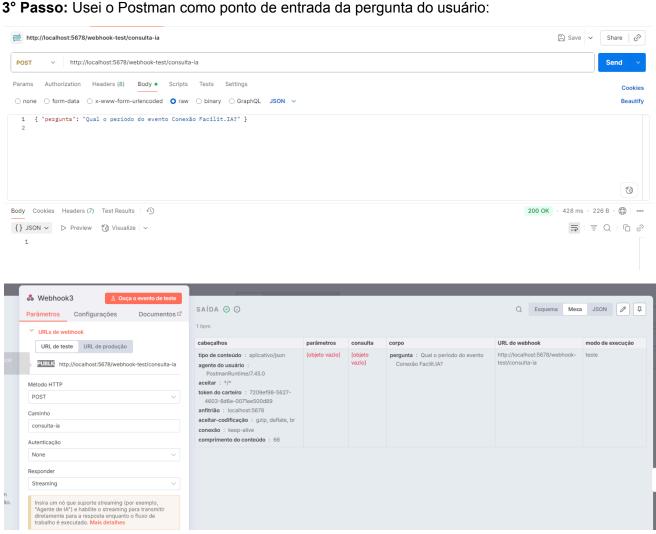
1° Passo: Acessei o site <a href="https://webhook.site/">https://webhook.site/</a> e automaticamente foi gerada uma URL exclusiva. Copiei essa URL e usei em um sistema que envia requisições HTTP. Assim que os dados foram enviados, visualizei tudo na tela.

https://webhook.site/#!/view/c09e4dff-7a8b-4aab-9f0f-74bc62ab58af/63c90750-2510-403e-aa20-bca45ecddee0/1

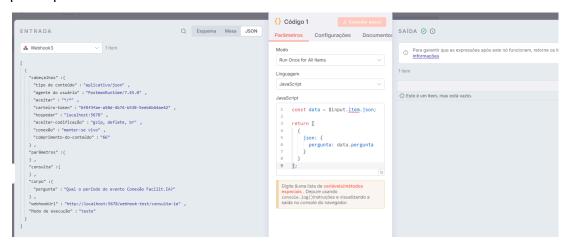


**2° Passo:** Criei um novo workflow no n8n e comecei adicionando uma **URL** teste: https://webhook.site/c09e4dff-7a8b-4aab-9f0f-74bc62ab58af



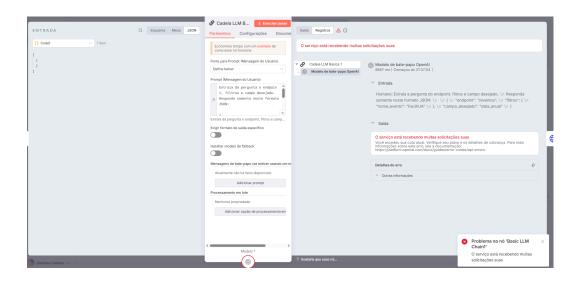


**4° Passo:** Adicionei um nó **code** apenas para extrair o campo pergunta da requisição e passa para frente no formato **JSON**:



**5° Passo:** Adicionei um nó **Basic LLM Chain1** (junto a OpenAl Chat Model), com a função de receber a pergunta e transformar em instruções estruturadas no formato **JSON**. Ou seja, ele entende qual **endpoint da API** chamar, quais filtros aplicar e qual campo buscar.

```
{
  "endpoint": "/eventos",
  "filtros": { "nome_evento": "Facilit.IA" },
  "campo_desejado": "data_anual"
}
```





(para adicionar na credencial do n8n, criei algumas chaves de testes, além da que foi disponibilizada no desafio)

**6° Passo:** Criei um nó **request** para montar a URL dinamicamente usando as variáveis do JSON do LLM:

http://api\_container:8000/eventos?nome\_evento=Facilit.IA



**7° Passo:** Adicionei um novo nó **Basic LLM Chain2** (com outra OpenAl Chat Model) para receber a pergunta original e os dados crus da API. A ideia é usar o LLM como "pós-processador", gerando um JSON de resposta mais natural e direto.



**8° Passo:** Criei um novo **code** para implementar uma lógica extra de validação, onde ele lê o JSON retornado da API. A ideia é verificar os campos desejados e o nome do evento que o LLM extraiu. Com isso, vai procurar o evento na lista da API e retorna apenas o valor específico ou "Evento não encontrado".

```
Exemplo:
```

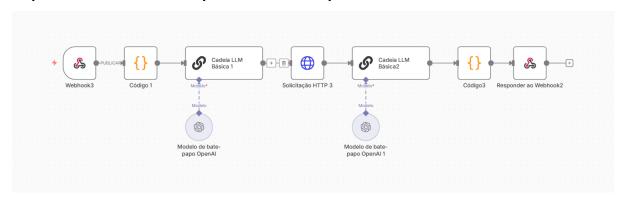
```
{
    "evento": "Facilit.IA",
    "campo": "data_anual",
```

```
"resposta": "01/05/2025 à 31/07/2025" }
```

**9° Passo:** O último nó adicionado foi o **Respond to Webhook**, retornando o JSON final ao cliente que chamou o webhook.

```
Exemplo de resposta:
{
    "evento": "Facilit.IA",
    "campo": "data_anual",
    "resposta": "01/05/2025 à 31/07/2025"
}
```

# Arquitetura do fluxo criado para a terceira etapa do desafio:



# **Desafios encontrados:**

Um dos maiores desafios que enfrentei neste projeto foi o primeiro contato com o n8n e a OpenAl. Nunca tinha trabalhado com automação de fluxo de dados ou com inteligência artificial, então tudo era novo e, muitas vezes, parecia um grande quebra-cabeça. A princípio, entender como o n8n conectava as planilhas do Google Sheets à minha API e depois ao banco de dados parecia bem complicado. A cada erro de JSON inválido ou de variáveis que não eram reconhecidas, a sensação de estar começando do zero voltava. No entanto, a cada problema resolvido, a cada nó que funcionava como o esperado, o processo se tornava incrivelmente gratificante. No final, percebi que por mais desafiador que tenha sido, o conhecimento que adquiri sobre ETL e integração de sistemas foi enorme. Foi uma jornada de aprendizado que me mostrou o poder da automação para simplificar tarefas complexas e que, com certeza, levarei para meus próximos projetos.